

ONLINE LOCALIZATION OF ZOONIVERSE CITIZEN SCIENCE PROJECTS – ON THE USE OF TRANSLATION PLATFORMS AS TOOLS FOR TRANSLATOR EDUCATION

by **Krzysztof Michalak**

Center for Theoretical Physics PAS

Al. Lotników 32/46, 02-668 Warsaw, Poland

krzysztof.michalak16 @ gmail.com

Abstract

This paper aims at describing the way in which online translation platforms can facilitate the process of training translators. *Zooniverse*, a website hosting a variety of citizen science projects in which everyone can take part, was used as an example of such a concept. The first section of this paper is focused on the history, idea and general description of the website. In section two the online translation platform, which is connected with *Zooniverse*, has been presented in detail with emphasis put on advantages and weak points. Ideas for practical application of *Zooniverse*'s localization platform have been outlined in section three.

Results have shown that localization platforms hold a great potential in terms of providing training for trainee translators. This is mostly because they offer basic experience in terms of simple computer assisted translation technologies and access to translations into other languages that have been already submitted. They are also characterized by simplicity and accessibility, as the platform can be used from any place all over the world. *Zooniverse* can, therefore, be the basis for further research on how the potential of such websites can be harnessed for more effective translator training.

1. Introduction

In the 21st century the Internet is the most popular of all media and the variety of its applications is simply incredible. It is, therefore, not surprising that the Internet is an extremely useful tool for scientific research and education. This can be illustrated perfectly by the example of the *Zooniverse* website. It is the mother of a great array of scientific projects focusing on biology, genetics, ecology, history and most of all astronomy and astrophysics. In this article I am going to present the idea behind the citizen science and the *Zooniverse* family of projects, as well as the whole website from the technical perspective. Also, I will describe the way the *Zooniverse* translation platform, which works with every *Zooniverse* project, can be applied for the purpose of training future translators.

1.1. Zooniverse in general

Zooniverse was created in July 2007 together with the onset of the first *Zooniverse* crowdsourcing project called Galaxy Zoo. Since its launching, many more projects have been created, with the majority of them enjoying unceasing and incredible success and popularity. The *Zooniverse* family of websites has been developed and maintained by the Citizen Science Alliance (CSA), which is run by seven different organizations and institutions from the UK and the USA, with the Center for Theoretical Physics PAS and New Science Foundation being CSA partners in Poland.

The *Zooniverse* citizen science projects are devoted to various areas of science. The website is composed of many webpages, each acting as a separate science project in which everyone can participate. All science teams involved require hundreds of thousands of images classified and categorized in order to make them useful. Those images come from different surveys and research carried out by the aforementioned science teams. The amount of visual data collected in every case is so large that the task of analyzing it tends to be impossible for those teams alone. Unfortunately, computers are not capable of classifying those images on their own either and due to this the human intervention tends to be unavoidable and much more accurate. This is where volunteers – groups of people who populate the Internet in abundance – come in!

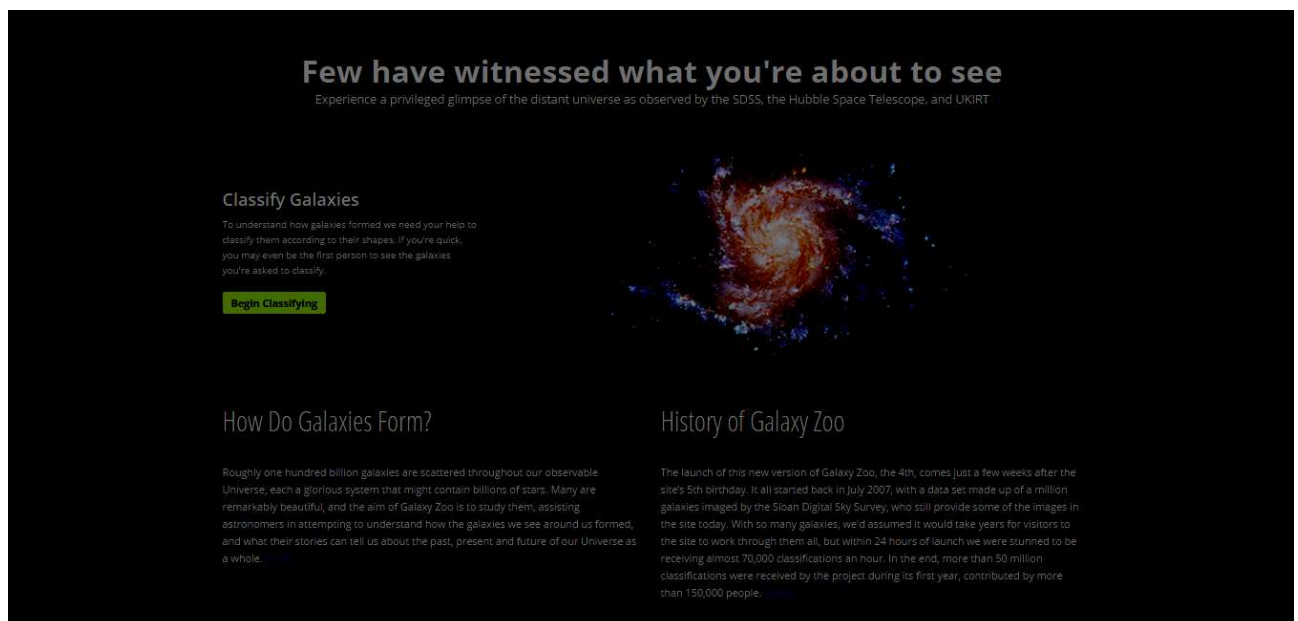


Figure 1. Galaxy Zoo homepage (source: www.galaxyzoo.org).

1.2. The idea behind citizen science

Citizen science is a new term which describes the concept of engaging the public in real scientific research and giving the researchers a hand in analyzing their enormous datasets so they can devote their precious time to more pressing matters. In order to become a citizen scientist one does not have to be a professional in any field. In fact, it is mainly addressed to people who have little to do with professional science on a daily basis.

The citizen science comes in a wide variety of forms. The most popular ones focus on providing support in data collection and in the analysis of the data already collected. *Zooniverse* takes advantage of the latter and gathers as many volunteers as possible with the aim of using their potential to enhance scientific works. It is a website with a global impact, so virtually anyone can sign in and help the scientists. By going through a short tutorial and then classifying just a couple of images, one can help make a real difference.

Furthermore, in order to increase the number of people involved in classifying and submitting the results a special translation platform was created for all projects. As a result, volunteering translators from all over the world can choose a project matching their interests or areas of expertise and translate it; consequently by doing so they expand the number of those taking part in advancing scientific research.

1.3. Operation and results

The way all the *Zooniverse* projects work is very simple. Signing up is not mandatory, but when an account is created by a potential user, it lets them see the classifications to be submitted, discuss the images with other volunteers and members of the science teams, share the results and take advantage of some other additional features. The first thing one visiting any *Zooniverse* website should do is to go through the brief tutorial and see the “about” section of the page in order to understand what the aims of the specific project are. Tutorials are created in a very user-friendly manner so there is no doubt what is expected from those who intend to join in the cooperation. The fact that the majority of the websites are translated into other languages makes them even more accessible. After completing the tutorial, a user is ready to classify the visual data available on the website and submit the results directly to the science team.

As every project requires volunteers to perform different tasks, the *Zooniverse* projects pose a wide range of opportunities for cooperation. Project types include annotation, decision tree, pattern matching, ranking, filtering and transcription.

Annotation requires volunteers to use a special tool to draw shapes on the image in order to mark specific features visible on the picture. In a decision tree project participants are asked a series of questions and are supposed to choose the answer which best reflects what the image shows. Pattern matching is based on categorizing the sounds or features of images according to specific examples, while in the ranking project one has to decide which image fulfills some requirements to a greater extent. Filtering projects are based on a gradual description of the visual data by analyzing the images and answering questions. Finally, transcription involves the process of deciphering the characters from an image in order to present them, as a result, in a typed form.

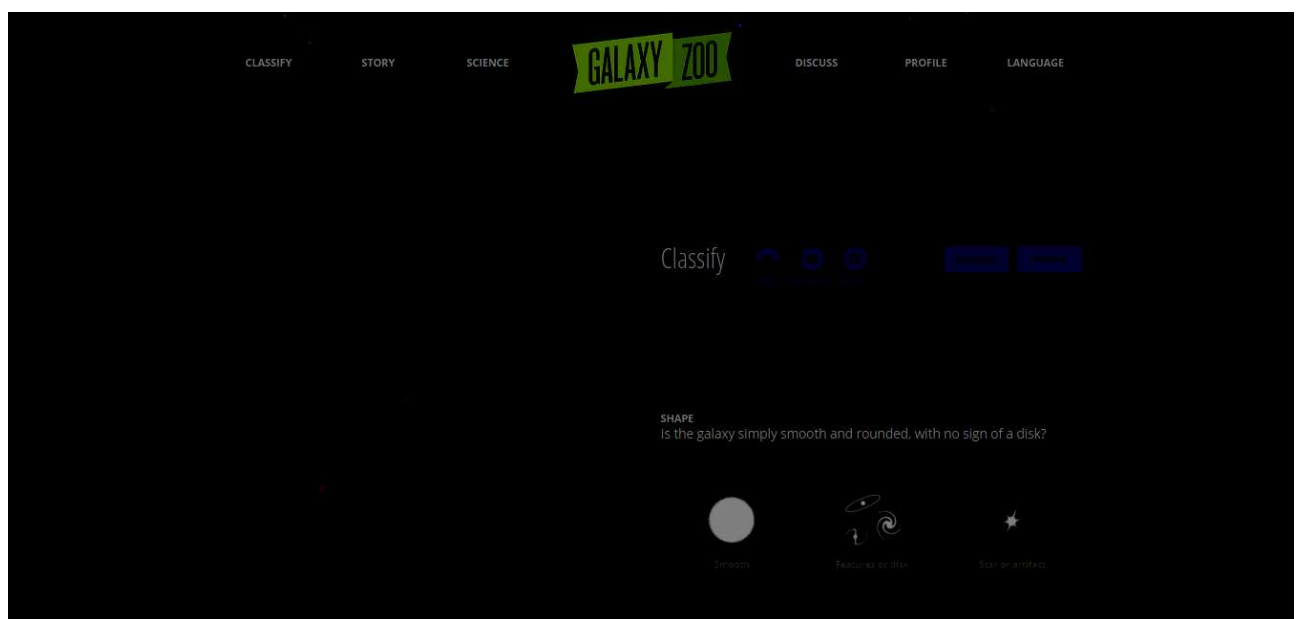


Figure 2. Galaxy Zoo classification applet (source: <http://www.galaxyzoo.org>).

All the data gathered from the projects are first stored and then analyzed by the researchers. It is quite plausible that the very data volunteers are helping to collect now will help scientists teach computers how to do the same job in the future, thanks to extensive sample data which can be used to feed computers for further analysis. Some of the data, like in the case of the *Zooniverse Galaxy Zoo* project, are published and can be accessed virtually by practically anyone all over the world. Moreover, as there are many volunteers classifying the images and each image is analyzed several times by different participants, very interesting discoveries tend to be made as a result of such a multi-angle approach. For example, thanks to the countless classifications collected by the Galaxy Zoo, scientists were able to discover such objects as “green peas”, being intensely star-forming galaxies or “Hanny's Voorwerp”, which is a very rare astronomical object.

1.4. Polish versions

A great majority of *Zooniverse* citizen science projects have been translated into many languages, including Polish. This paper was prepared on the basis of the translation work carried out for the sake of creating Polish versions of available projects in order to make them more accessible to non-English speaking Polish audience.

Volunteers from Poland managed to make a modest but still valuable contribution to the operation of *Zooniverse* projects like for instance Galaxy Zoo, Radio Galaxy Zoo, Disk Detective and Milky Way Project. At the turn of the year 2014 Polish version of Galaxy Zoo took the 6th place, Radio Galaxy Zoo and Disk Detective were 5th and Milky Way Project was 7th among all other language versions according to the number of sessions. Some of the Polish versions have been around only for a few months, so they are still gaining their popularity. The data presented show that translation of *Zooniverse* projects was not in vain and helped make a difference.

2. Description of the translation platform

Having outlined what *Zooniverse* is and what it aims at, I will describe the translation platform itself together with all its features. First I will go through the technicalities of the site, briefly describing how the platform works and what it offers. Then I will delve deeper into what advantages and disadvantages of this translation platform for its potential users.

2.1. Technical aspects

As I have previously mentioned, every *Zooniverse* citizen science project is connected to a special translation platform which supports all projects collectively. In order to use it one has to create an account by registering via any of the available projects and then be granted access to the platform by one of the *Zooniverse* crew members. After choosing a specific project and being assigned to it, one can begin the process of translation into a language that is at the time unavailable for a given project. The whole process ensures that only selected, authorized users have access to the translated material; this policy prevents any translation from being submitted, deleted or even modified accidentally.

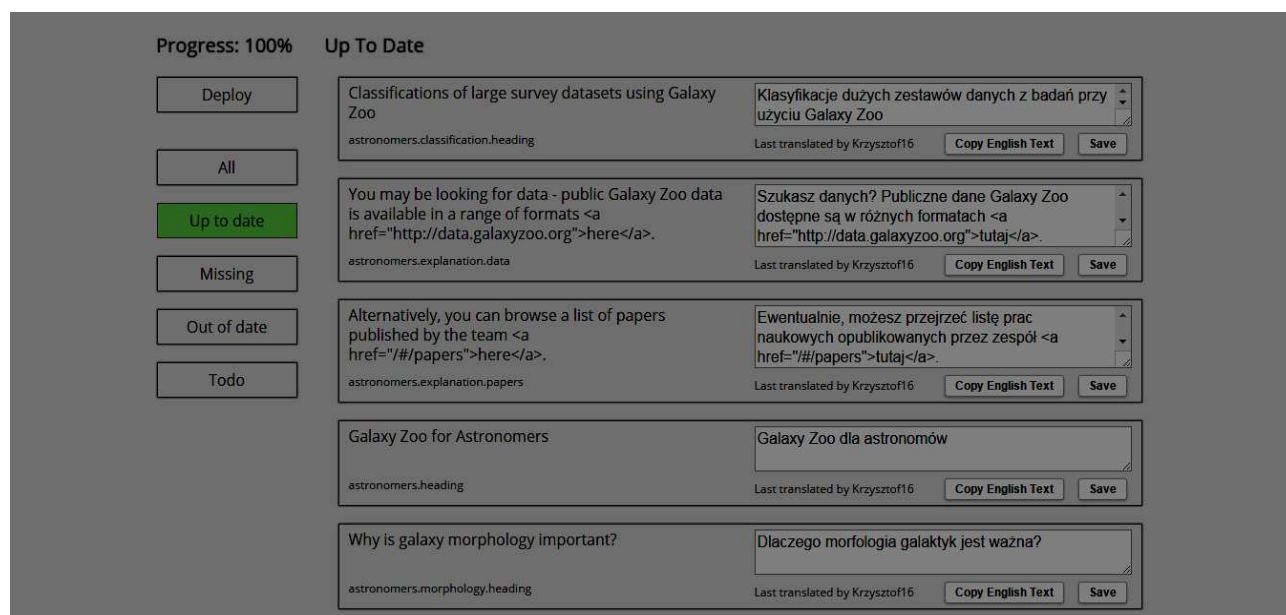


Figure 3. An excerpt from the *Zooniverse* translation platform (source: <http://translations.zooniverse.org>).

The platform offers a few options for the user. The text of a website is divided into many windows, each of which represents single words, chunks of words, sentences or sometimes multiple paragraphs. Next to every window containing the text in the source language (SL) there is a blank window where the text in the target language (TL) should be typed. When the text from a window is translated, it can be saved by simply clicking a button. Of course the already translated and submitted material is still available for any further modification or proofreading. When the text from every window is translated and checked, the translator can contact one of the members of the *Zooniverse* team in order to publish the translated version of the site for public use.

Further on, this paper aims at describing the way the translation platforms, as the one described above, can be used to train translators. Therefore, I will now try to explain what are the advantages and disadvantages of such translation platforms based on the example of the one created by *Zooniverse*.

2.2. Advantages

The first advantage I would like to mention is the clarity of the platform. The fact that each paragraph and sometimes single words or sentences are presented in a separate window is a feature that translators may welcome as a very convenient option. As a result, a limited chunk of the source text is clearly presented for the translation process. What is more, after the translation the text in the SL and in the TL are right next to each other and are clearly available for any possible editing processes. This accounts for the ease of translation and

prevents chaos. This is also a perfect introduction to the concept of treating a text as an entity divided into segments for the translation purpose, which is a well-known concept of the CAT technology. Moreover, the already translated text is saved in different tabs, including “Up to date”, “Out of date”, “Missing”, “To do” or “All”. As a result, the translator can easily navigate between the translated and non-translated material.

Secondly, the platform is equipped with a very useful function that can be of great advantage for both aspiring and professional translators. Namely, it allows generating a trial test version of the website with the translated content that is available only to the translator. This makes proofreading much easier and more comfortable, as one can see the site exactly the way it would look like with the currently translated text. Thanks to that, a translator can see whether he or she has committed any mistakes in terms of paragraph division, spelling and transfer or detect any errors in the HTML code. After spotting a mistake, one simply needs to find the respective window on the platform, type the corrected text or code, save the window and redeploy the whole site.

Another advantage is the fact that any translator of a particular language pair that has access to a given project can browse through other language versions of the same project. As most of the *Zooniverse* projects are translated into several languages, this is an incredibly useful feature, especially for translators who know more than one foreign language. Whenever one encounters a translation problem caused by insufficient context, unclear, ambiguous and possibly completely unknown vocabulary or significant mistakes in the source text, it is possible for the translator to make use of the translations submitted by translators of other languages. These translations show how other translators solved the problem or translated a particular item into their first language. Sometimes the translation submitted by one translator can be a combination of translations or strategies that were adopted by more than one translator; therefore, it is a very useful functionality, as it can offer the translator various translation solutions.

There is also one advantage that is of lesser significance for the translator than those listed above; however, it is still worth mentioning. Because the content of the website is presented to the translator as a lengthy list of windows with corresponding pieces of text, the platform has a counter that shows the percentage of windows already completed. Although this feature is a bit imperfect, as it does not take into account the length of the segments but their number only, it enables the translator to gain an overall impression of how much work there is left and potentially how the translation process can be made more efficient.

2.3. Disadvantages

The *Zooniverse* translation platform also has some disadvantages, which are present due to the fact that it is actually a simple tool, not professional translation online software. The flaws that can be encountered while performing translation tasks on the platform are not too frequent. However, whenever the translator faces them, he or she is bound to spend more time than necessary in order to solve the problem. I believe some of the issues can be easily fixed or patched, but it is possible that in order to fix some of the other problems, serious restructuring of the whole translation platform is required. I am going to go through only some of the problems, those that are most common.

One of the problems a translator working on such a translation platform is surely going to encounter is the presence of incredibly lengthy pieces of text represented by a single window. In such a case one has to copy the whole text into some other text processing program, as this long text is very inconveniently presented. What is more, the text is often interwoven with the HTML code, which makes the situation even more complicated (the issue of the HTML code will be addressed in the next paragraph). This problem can be eliminated by the technical staff of the website whose task is to properly divide the content into windows representing specific pieces of text on the platform. Presenting the SL material in such a form is highly undesirable for the translator and significantly extends the time it takes to complete the translation process making it a significantly more challenging task. Consequently, this increases the number of translation errors, committed in particular by inexperienced translators.

Another significant drawback of this translation platform is the aforementioned problem of the HTML code being interwoven with the text itself. Therefore, a translator taking up the translation has to be familiar at least with the most basic HTML tags. As sometimes lines of code are an integral part of the text, they have to be copied exactly in the same form from the SL text into the TL text. Any misspelling of the tags may result in errors in the display of a particular fragment of text on the website, which may result in it becoming completely incomprehensible. Therefore, it is not only problematic, but also highly time-consuming. What makes things even worse is the fact that there is no way to input these tags by other means than typing, or copying and pasting. The platform, unfortunately, lacks a panel with basic HTML tags, so whenever they appear in the text, they need to be typed, or copied and pasted by the translator from the source text. Sometimes one can get lost in the abundance of tags and symbols which can lead to unpredictable outcomes.

There are also a few other problems which, however, are far from being a great

inconvenience. One of them is the lack of the option of any basic text statistics, like a word count for example. Of course, it is regarded as a disadvantage only from the translator's point of view, since the purpose of this paper is to evaluate to what extent this platform may facilitate and simplify the translation process. As this form of translation could be used as a code of good practice, even for academic purposes, the possibility to quickly count words of the translated text is highly desirable, although this feature shall only be valued from the translation process perspective. However, the fact that the information about the progress in translation (the percentage of text translated described in section 2.2) is very general might turn out to be slightly misleading for the translator. It is because the counter does not take the window size into account. Lastly, one can sometimes encounter a bug while generating a test version of the site. This bug prevents some paragraphs from being exported and as a result they still appear in English, not in the TL, even though they have been properly translated and saved. In such a case, the only option the translator is left with to solve such problems is to contact the technical support team.

3. Practical applications

The very purpose of this paper is to show that translation platforms, such as the *Zooniverse* platform described above, hold educational potential as a practical activity for in-training and aspiring translators. The fact that translations done via such platforms are online and can be performed from any location makes it very easy to use for education. I will now outline how translation platforms could be utilized by teachers and lecturers for providing translation practice to their students and what are the aspects of such platforms that emphasize its educational potential.

Translations done via the platform could easily become an assignment for students to complete either individually, in pairs or in larger project-based groups. The platform facilitates collective translation by showing which particular translator is responsible for translating specific parts of the website. Therefore, after the assigned translation is completed and submitted for review and grading, the person responsible for evaluation can see which segments have been translated by particular students and grade them accordingly.

The teacher can access the translation platform and review the segments together with their corresponding windows. Each segment is signed with the username of each particular translator who has submitted the latest version of the translation, which aids the process of grading performed by the reviewer. Moreover, the teacher responsible for assessing the translation can use the export function in order to view the test version of the site and see how

the proper text is composed.

A great training potential also lies in the features and possibilities the platform has to offer, as described above. The ability to confront one's translation with those submitted in different foreign languages can point at a variety of other strategies and solutions that can be applied in the translation process. It is also a significant advantage when it comes to providing more contexts for the purpose of decoding the specific information hidden in the text. Also, as mentioned before, such assignments introduce aspiring translators to the concept of dividing texts into segments for translation purposes, which is very similar to the concept offered by any CAT tool.

4. Summary

The advantages and various aspects of translation platforms, such as the one of *Zooniverse*, prove their educational potential in providing hands-on translation training for beginner translators. They are confronted with a wide range of possibilities in terms of comparative analysis of translations submitted by them and fellow volunteers. The platform is characterized by its high transparency and relative ease of operation. Although it is not free of flaws and disadvantages, I believe it is possible to improve and reprogram it to suit the translators' needs in order to raise the quality of translations. Lastly, it is worth taking into account that the translation platform has not been created for any educational purposes in particular. If the platform had been created from scratch to meet such objectives, it would, undoubtedly, have even higher training potential. This suggests, therefore, the need for further research into this matter.